AD_____

Award Number: DAMD17-03-1-0475

TITLE: Integrating Geographic Information System (GIS) Into Breast Cancer Epidemiologic Research

PRINCIPAL INVESTIGATOR: Daikwon Han, Ph.D.

CONTRACTING ORGANIZATION: New York State University Research Foundation
Amherst, New York 14228

REPORT DATE: July 2004

TYPE OF REPORT: Annual Summary

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	July 2004	Annual Summary (1 Jul 2003 - 30 Jun 2004)

4. TITLE AND SUBTITLE
Integrating Geographic Information System (GIS) Into Breast Cancer Epidemiologic Research

5. FUNDING NUMBERS
DAMD17-03-1-0475

6. AUTHOR(S)
Daikwon Han, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
New York State University Research Foundation
Amherst, New York 14228

E-Mail: dhan@buffalo.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. Abstract *(Maximum 200 Words)* *(abstract should contain no proprietary or confidential information)*

The objective of this postdoctoral training research is an integration of GIS and a spatio-temporal perspective into breast cancer research of the relationship between environmental exposures and breast cancer risk. As proposed, developing a plan for modeling and development of the data were a major component of the first year goals. For this goal, I have been focusing on completion of spatial clustering analyses of residences. Also we completed updating of the lifetime residential histories for our dataset for the breast cancer cases and controls. We have completed collection of historic traffic information, and a GIS-based traffic model was established. These data will now be used in my future analyses. The development of a theoretical framework measuring similarity and difference of individual's lifetime residential history is in progress. Training in epidemiology was another major task. I have been involved in analysis and writing of a classic epidemiologic research paper and have participated in several workshops as a part of epidemiology training.

14. SUBJECT TERMS
environmental epidemiology

15. NUMBER OF PAGES
24

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

Table of Contents

INTRODUCTION

The objective of this postdoctoral training research is an integration of Geographic Information Systems (GIS) and a spatio-temporal perspective into breast cancer research of the relationship between environmental exposures and breast cancer risk. With the increased use of GIS in epidemiologic studies, it becomes possible to examine lifetime exposures to environmental risk factors by integrating lifetime exposure information in GIS with other breast cancer epidemiologic factors. As a part of the on-going case-control study of breast cancer in western New York, the proposed study examines the relationships between the residential environment and breast cancer risk and test the hypothesis, 1) Lifetime cumulative exposure to Polycyclic Aromatic Hydrocarbons (PAHs) and benzene will be more strongly associated with risk for breast cancer than of any one time period. 2) There will be sensitive periods in a woman's life that will carry greater risk for exposure. Specific aims of the proposed study include, 1) Continuing evaluation on the role of environmental risk factors on breast cancer; we will develop a GIS-based model of lifetime residential history and environmental exposure in breast cancer. 2) Assessment of historical exposures to PAHs and benzene and breast cancer risk; we will reconstruct historical exposure to PAH and benzene for the use in environmental epidemiology of breast cancer and empirically assess exposure to these two environmental compounds based on lifetime exposure index. Using lifetime residential information for breast cancer cases and controls in western New York, the proposed study examines breast cancer risk from lifetime exposures. Lifetime residential histories have been collected for the breast cancer cases and controls by interview, while information on environmental contaminants is being collected from historical sources. We will use GIS in our assessment of the associations of environmental risk factors and breast cancer incidence. Further spatial-statistical analysis will be performed in a GIS environment to examine associations between residential environment and breast cancer risk in spatial and temporal dimensions. There are no major results at this time, but some preliminary findings are discussed in the text of this report.

BODY OF REPORT

Task 1: Developing plan for modeling, data needs, and training.

Task 1 is completed. As proposed, developing a plan for modeling and development of the data were a major component of the first year goals. For this goal, I have been focusing on completion of spatial clustering analyses of residences.

I completed a GIS-based spatial and temporal analysis for residences of breast cancer cases and controls at early life and found strong evidence of spatial clustering for cases during this time. A paper on geographic clustering of residence in early life and subsequent risk of breast cancer has been accepted for publication (Han, D. Rogerson, PA. Nie, J. Bonner, MR. Vena, JE. Muti, P. Trevisan, M. Edge, S. Freudenheim, JL. 2004. Geographic Clustering of Residence in Early Life and Subsequent Risk of Breast Cancer. 2004. Cancer Causes and Controls 15: 921-929). My paper based on this work was selected as one of ten finalists for the Nystrom competition of the Association of American Geographers (AAG), and the paper was presented at the centennial meeting of the AAG, Philadelphia, PA. March 15, 2004. A copy of the abstract is in Reportable Outcome Section, and a copy of the manuscript is included in Appendix.

We also completed updating of the lifetime residential histories for our dataset for the breast cancer cases and controls. All Erie and Niagara county residential location were identified and geocoded, and these were merged into one database. We checked consistency of geocoded addresses in different time points for each individual, updated incomplete addresses using Polk searches, and validated the consistency of reported years of moved in and out of the residence. A manuscript is in preparation regarding the clustering of lifetime residence and breast cancer risk using exploratory spatial analysis tools based on these lifetime residential history data. An abstract based on this work was presented at the annual meeting of the International Society for Environmental Epidemiology in New York City, New York, August, 2004. A copy of the abstract is in Reportable Outcome Section.

Second, we have now completed collection of the PAHs and benzene exposure from historic traffic information and air pollution sources. Working with my colleagues Dr. Jing Nie and Dr. Matthew Bonner and others, we have assessed historical exposure to PAH from these sources, and found evidence of association between PAHs exposures from traffic and air pollution sources in relation to breast cancer risk, especially PAH exposures during sensitive time periods in early life. Also a GIS-based traffic model was established to estimate historical residential exposure to PAHs from traffic, and a geostatistical method was utilized to predict and interpolate individual residential TSP concentration for the estimation of PAHs exposure from air pollution source. These data and evidence from epidemiologic studies will be used in my future spatial analyses of lifetime exposure to PAHs and breast cancer risk.

Training in epidemiology was another major task. I have been involved in analysis and writing of a classic epidemiologic research paper and have participated in several workshops as a part of epidemiology training. As a part of training, I took an epidemiologic methods course, "SPM 502: Advanced Epidemiologic Methods" in the Department of Social and Preventive Medicine, University at Buffalo. I participated in the NCI Summer Curriculum in Cancer Prevention, and took the course "Principles and

Practice of Cancer Prevention and Control" in July, 2003. I was selected as one of the participants in Center for Spatially Integrated Social Science (CSISS) summer workshop, "Geographically weighted regression and associated statistics" in August, 2003.

Finally, in order to increase my skills in epidemiologic analysis, I have been working on a more classic exposure-disease analysis of lifetime body weight and breast cancer risk. The paper is however a departure from the classical analysis in that I am looking at early exposures and modeling cumulative exposures as well as exposures in potentially sensitive time periods. I have been using the same breast cancer study as the one for my GIS research, a case-control study in western New York, the Western New York Exposures and Breast Cancer (WEB) study. An abstract based on this work was accepted to the annual meeting of the Society for Epidemiologic Research in Salt Lake City, and a poster based on this abstract was presented on June 15, 2004. A manuscript is in preparation regarding effects of lifetime adult weight change on pre- and post-menopausal breast cancer risk using breast cancer case-control data in western New York.

Task 2: Developing and testing a model for lifetime exposure and breast cancer risk

Task 2 is underway. The development of a theoretical framework measuring similarity and difference of individual's lifetime residential history is in progress. Since investigators in the Department of Geography and NCGIA have developed GIS-based theoretical frameworks measuring similarity or dissimilarity of individual's lifetime residential history in space and time, I will apply models of geospatial lifeline to breast cancer case and control data in western New York.

Task 3: Assessing historical exposures to PAHs and benzene and breast cancer risk.

We have not begun estimating historic exposure to PAHs and benzene. However, we have collected and analyzed the PAHs and benzene exposure information from lifetime smoking histories (active and passive smoking histories), traffic roadways, air pollution sources (total suspended particulates), and industrial sites. These data will now be used in my future spatial analyses, based on the models of geospatial lifeline for the estimation of lifetime residential exposures to PAHs and benzene and breast cancer risk.

KEY RESEARCH ACCOMPLISHMENTS

- I completed a GIS-based spatial and temporal analysis for residences of breast cancer cases and controls at early life and found strong evidence of spatial clustering for cases during this time. A paper on geographic clustering of residence in early life and subsequent risk of breast cancer has been accepted for publication. (Han, D. Rogerson, PA. Nie, J. Bonner, MR. Vena, JE. Muti, P. Trevisan, M. Edge, S. Freudenheim, JL. 2004. Geographic Clustering of Residence in Early Life and Subsequent Risk of Breast Cancer. Cancer Causes and Controls 15:921-929)
- We completed updating of lifetime residential history of breast cancer cases and controls in western New York; All Erie and Niagara county residential location were identified and geocoded, and these were merged into one database. We checked consistency of geocoded addresses in different time points for each individual, updated incomplete addresses using Polk index searches, and validated the consistency of reported years of moved in and out of the residence.
- A manuscript is in preparation regarding the clustering of lifetime residence and breast cancer risk using exploratory spatial analysis tools based on these lifetime residential history data.
- We also completed collection of the PAHs and benzene exposure from historic traffic information and air pollution sources, and a GIS-based traffic model was established to estimate historical residential exposure to PAHs from traffic.
- An abstract was presented at the annual meeting of the Society for Epidemiologic Research in Salt Lake City, Utah, June 2004. (American Journal of Epidemiology Supplement 159: S13)
- An abstract was presented at the annual meeting of the International Society for Environmental Epidemiology in New York City, New York, August, 2004. (Epidemiology Supplement 15: S529)
- My paper was selected as one of ten finalists for the Nystrom competition of the Association of American Geographers (AAG), and the paper was presented at the centennial meeting of the AAG, Philadelphia, PA. March, 2004.
- I gave an invited talk on "GIS, Medical geography, and Spatial Epidemiology" in Kyung Hee University, Seoul, Korea. June, 2004.
- I presented a seminar, "Geographical Epidemiology of Breast Cancer in Western New York" as part of the seminar series of the Department of Social and Preventive Medicine, University at Buffalo. March, 2004
- As a part of training, I took an epidemiologic method course, "SPM 502: Advanced Epidemiologic Methods" in the Department of Social and Preventive Medicine, University at Buffalo. Spring, 2004, 3 credit hours.
- I participated in the NCI Summer Curriculum in Cancer Prevention, and took the course "Principles and Practice of Cancer Prevention and Control" in July, 2003.
- I was selected as one of the participants in Center for Spatially Integrated Social Science (CSISS) summer workshop, "Geographically weighted regression and associated statistics" in August, 2003.

REPORTABLE OUTCOMES

Abstracts/Presentations

- "Effects of Lifetime Weight Gain on Breast Cancer Risk" abstract presented at the annual meeting of the Society for Epidemiologic Research, Salt Lake City, UT. June 2004. (abstract attached)
- "Assessing the Variability of Risk Surfaces using Residential History Data in a Case Control Study of Breast Cancer" abstract accepted to the annual meeting of the International Society for Environmental Epidemiology, New York, NY. (abstract attached)
- "Geographic Clustering of Residence in Early Life and Risk of Breast Cancer" paper presented at the Annual Meeting of the Association of American Geographers, Philadelphia, PA. March 2004. (abstract attached)

Publications

- Daikwon Han, Peter Rogerson, Jing Nie, Matthew Bonner, John Vena, Dominica Vito, Paola Muti, Maurizio Trevisan, Stephen Edge, Jo Freudenheim. 2004, "Geographic Clustering of Residence in Early Life and Subsequent Risk of Breast Cancer" Cancer Causes and Control 15: 921-929. (manuscript attached)

Effects of Lifetime Weight Gain on Breast Cancer Risk D Han, P Muti, M Trevisan, J Nie, D Vito, S Edge, J Freudenheim. Annual Meeting of the Society for Epidemiologic Research, Salt Lake City, UT. 2004.

While there is quite consistent data regarding increased risk of postmenopausal breast cancer with increased body mass index, there is now accumulating data that would indicate that weight gain in adult life is more predictive of breast cancer risk. In this study, we investigated effects of lifetime weight gain on pre- and postmenopausal breast cancer, and effects of weight changes at specific time points in a woman's life. A population-based case control study, the Western New York Exposures and Breast Cancer Study(the WEB study) was conducted. Included were 1,170 women with primary, histologically confirmed, incident breast cancer and 2,116 controls frequency-matched on age and race. Participants were asked to recall their body weight for each decade of their lives from age 20 to the present. Total lifetime weight gain, the difference between weight one year before interview and weight at age 20, and weight changes between each decade were examined. Unconditional logistic regression was used to estimate odds ratios (OR) and 95% confidence intervals (CI). An increased risk of breast cancer was found for postmenopausal women in the highest (> 27 kg) compared to the lowest quartile (< 9.9 kg) of total lifetime weight gain (adjusted OR 3.28, 95% CI 2.16-4.98). Risk estimates increased in magnitude for increasing time periods of weight gain. Among the premenopausal women, lifetime weight gain was not associated with an increased risk of breast cancer. This study confirms previous findings of increased risk of breast cancer associated with adult weight gain among post- but not premenopausal women, and suggest that weight gain over longer periods of time is associated with higher risk.

Assessing the Variability of Risk Surfaces using Residential History Data in a Case Control Study of Breast Cancer. Daikwon Han, Peter A Rogerson, Matthew R Bonner, Jing Nie, John E Vena, Jo L Freudenheim. Annual meeting of the International Society for Environmental Epidemiology, New York, NY. 2004

Introduction: Residential location has often been used as a measure of environmental exposure in epidemiologic studies. To examine breast cancer risk associated with residential history, we explored the spatio-temporal patterns of risk surfaces using data on lifetime residential history in a case control study of breast cancer. We applied GIS-based exploratory spatial analyses to obtain risk surfaces, and assessed spatio-temporal variability of the risk surfaces.
Methods: A population-based case control study of breast cancer in western New York (the WEB Study) includes data on the lifetime residential history for breast cancer cases and controls. Participants were asked to provide all locations of earlier residences. Density surfaces of cases and controls were obtained using kernel estimation methods, and the standardized difference in density surfaces was identified to depict elevated areas of breast cancer risk. The significance of the resulting risk surfaces was tested and reported as *p*-values. These surfaces were compared for premenopausal and postmenopausal women. To assess the variability of risk surfaces in space and time, the standardized difference in density surfaces was obtained for specific time periods in a woman's life and for each decade, from 1940s to 1990s.
Results: We found strong evidence of clustering of lifetime residence for premenopausal women (for cases relative to controls), and little evidence of such clustering for postmenopausal women. We also identified the time points contributing most significantly to this result. When density surfaces between cases and controls were compared at each time point, we observed that the earlier decades and early time points, such as residence at birth and menarche, were more likely to be influential time points in understanding overall patterns, relative to the importance of later time points.
Discussion: We were able to pinpoint geographic areas with higher risk, and to assess temporal variability of the risk surfaces by identifying the role of early exposures through exploratory spatial analyses.

Geographic Clustering of Residence in Early Life and Risk of Breast Cancer. Daikwon Han, Annual meeting of the Association of American Geographers, Philadelphia, PA. 2004.

This study examines breast cancer risk associated with lifetime residential history using GIS-based exploratory spatial analyses. Data on residential history and risk factors were collected as part of a population-based case control study of breast cancer in western New York. We were interested in whether there was clustering of breast cancer based on residential location in early life. Under the hypothesis that early exposures may be related to risk of breast cancer, *k*-function differences between breast cancer cases and controls were obtained. We found a general tendency of spatial clustering for residence in early life, compared with the simulated theoretical distribution of expected patterns. The evidence for clustered residential location at birth and at menarche was stronger than that for first birth or other time periods in adult life. Second, relative risk surfaces of cases and controls were obtained to depict elevated areas of breast cancer risk using kernel smoothing methods. We observed stronger evidence of geographic clustering of lifetime residence for pre-menopausal women relative to that for post-menopausal women, and clustering of early-life residence relative to that of adult-life residence. Our findings suggest that there may be identifiable etiological processes linking exposure and breast cancer risk, especially for pre-menopausal women, and that early exposures may be of particular importance. This study provides additional evidence that early environmental exposures may be related to breast cancer risk.

CONCLUSIONS

There are no conclusions regarding the major hypothesis to report at this time.

REFERENCES

1. Wolff, MS. Collman, GW. Barrett, C. and Huff, J. Breast cancer and environmental risk factors: epidemiologic and experimental findings. *Annual Review of Pharmacology and Toxicology* 1996; 36:573-596.
2. Laden F. and Hunter DJ. Environmental risk factors and female breast cancer. *Annual Review of Public Health* 1998; 19:101-123.
3. Kuh, D. Ben-Shlomo, Y. *A life course approach to chronic disease epidemiology: tracing the origins of ill-health from early to later life.* Oxford: Oxford University Press; 1997.
4. Barker, DJP. *Fetal and infant origins of adult disease.* London: BMJ Publishing; 1992.
5. Colditz, GA. Frazier, AL. Models of breast cancer show that risk is set by events of early life: prevention efforts must shift focus. *Cancer Epidemiology, Biomarkers, and Prevention* 1995; 4:567-571.
6. Trichopoulos, D. Hypothesis: does breast cancer originate *in-utero*? *Lancet* 1990; 335:939-940.
7. Potischman, N. and Troisi, R. *In-utero* and early life exposures in relation to risk of breast cancer. *Cancer Causes and Control* 1999; 10:561-573.

8. Elliott, P., Wakefield, J.C., Best, N.G., and Briggs, D.J. (2000). *Spatial epidemiology: methods and applications.* Oxford: Oxford University Press.

9. Ripley, BD. (1981) *Spatial statistics*. New York: Wiley.

10. Cressie, N. (1991) *Statistics for spatial data.* New York: Wiley.

11. Bailey TC, & Gatrell AC. (1995). *Interactive Spatial Data Analysis*. Harlow, England: Longman Scientific & Technical.

12. Bowman AW, & Azzalini A. (1997). *Applied Smoothing Techniques for Data Analysis.* Oxford: Clarendon Press.

13. Brody JG, Rudel RA, Maxwell NI, & Swedis SR. (1996). Mapping out a search for environmental causes of breast cancer. *Public Health Report* 3, 495-507.

14. Brody JG, & Rudel RA. (2003). Environmental pollutants and breast cancer. *Environmental Health Perspectives* 111, 1007-1019.

15. Cromley EK, & McLafferty SL. (2002). *GIS and Public Health.* New York: The Guilford Press.

16. DeBruin LS, & Josephy PD. (2002). Perspectives on the chemical etiology of breast cancer. *Environmental Health Perspectives Supplement* 110, 119-128.

17. Diggle PJ. (1983). *Statistical Analysis of Spatial Point Patterns*. London: Academic Press.

18. Khan OA. (2003). *Geographic Information Systems and Health Applications*. Hershey, PA: Idea Group Publishing.

19. Lawson A, Biggeri A, Bohning D, Lesaffre E, Viel J, & Bertollini R. (1999). *Disease Mapping and Risk Assessment for Public Health*. Chichester: John Wiley and Sons.

20. Lewis-Michl EL, Melius JM, Kallenbach LR, Ju CL, Talbot TO, et al. (1996). Breast cancer risk and residence near industry or traffic in Nassau and Suffolk Counties, Long Island, New York. *Archives of Environmental Health* 51, 255-65.

21. Paulu C, Aschengrau A, & Ozonoff D. (2002). Exploring associations between residential location and breast cancer incidence in a case-control study. *Environmental Health Perspectives* 110, 471-478.

22. Sabel CE, Gatrell AC, Loytonen M, Maasilta P, & Jokelainen M. (2000). Modelling exposure opportunities: estimating relative risk for motor neurone disease in Finland. *Social Science and Medicine* 50, 1121-37.

23. Silverman BW. (1986). *Density Estimation for Statistics and Data Analysis*. London: Chapman and Hall.

24. Timander LM, & McLafferty SL. (1998). Breast cancer in West Islip, NY: a spatial clustering analysis with covariates. *Social Science and Medicine* 46, 1623-1636.

25. Venables WN, & Ripley BD. (1998). *Modern Applied Statistics with S-Plus*. New York : Springer.

26. Wartenberg D. (2001). Investigating disease clusters: why, when and how? *Journal of the Royal Statistical Society A* 164, 13-22

APPENDIX

Cancer Causes and Control **15**: 921–929, 2004.

Geographic clustering of residence in early life and subsequent risk of breast cancer (United States)

Daikwon Han[1,2,*], Peter A. Rogerson[2,3], Jing Nie[1], Matthew R. Bonner[1], John E. Vena[4], Dominica Vito[1], Paola Muti[1], Maurizio Trevisan[1], Stephen B. Edge[5] & Jo L. Freudenheim[1]

[1]*Department of Social and Preventive Medicine, University at Buffalo, Buffalo, NY 14214, USA;* [2]*Department of Geography and National Center for Geographic Information and Analysis, University at Buffalo, Buffalo, NY 14261, USA;* [3]*Department of Biostatistics, University at Buffalo, Buffalo, NY 14214, USA;* [4]*Department of Epidemiology and Biostatistics, University of South Carolina, Columbia, SC 29208, USA;* [5]*Department of Breast and Soft Tissue Surgery, Roswell Park Cancer Institute, Buffalo, NY 14263, USA*

Received 23 March 2004; accepted in revised form 3 June 2004

Key words: breast cancer, early-life exposure, spatial clustering.

Abstract

Objective: This study focused on geographic clustering of breast cancer based on residence in early life and identified spatio-temporal clustering of cases and controls.
Methods: Data were drawn from the WEB study (Western New York Exposures and Breast Cancer Study), a population-based case–control study of incident, pathologically confirmed breast cancer (1996–2001) in Erie and Niagara counties. Controls were frequency-matched to cases on age, race, and county of residence. All cases and controls used in the study provided lifetime residential histories. The k-function difference between cases and controls was used to identify spatial clustering patterns of residence in early life.
Results: We found that the evidence for clustered residences at birth and at menarche was stronger than that for first birth or other time periods in adult life. Residences for pre-menopausal cases were more clustered than for controls at the time of birth and menarche. We also identified the size and geographic location of birth and menarche clusters in the study area, and found increased breast cancer risk for pre-menopausal women whose residence was within the cluster compared to those living elsewhere at the time of birth.
Conclusion: This study provides evidence that early environmental exposures may be related to breast cancer risk, especially for pre-menopausal women.

Introduction

Breast cancer is one of the leading causes of death among women in the United States. However, the epidemiology of breast cancer is not yet fully understood. We also do not fully understand mechanisms for the known risk factors; for instance, why changes in age at menarche or age at first birth have an impact on breast cancer risk. A substantial degree of geographical variation in breast cancer incidence and mortality in the US has been observed [1, 2]. While inconclusive, several environmental risk factors are also believed to be involved in breast cancer incidence [3, 4]. There is speculation that environmental factors may explain geographic variation in breast cancer rates not explained by known risk factors. For this reason, the potential role of environmental exposures in breast cancer risk is of particular interest.

In addition, there is a growing interest in early life and lifetime exposures in relation to breast cancer risk. The life course approach is of interest because there may be sensitive time periods for exposures and/or there may be cumulative effects of lifetime exposure involved in breast cancer incidence [5, 6]. Early life has an effect on breast cancer etiology evidenced by the known risk factors

*Address correspondence to: Daikwon Han, Department of Social and Preventive Medicine, University at Buffalo, 3435 Main St., Farber Hall, Rm 270, Buffalo, NY 14214, USA. Ph.: +1-716-829-2975 ext. 605; Fax: +1-716-829-2979; e-mail: dhan@buffalo.edu

such as age at menarche, age at first birth and parity. There is new evidence that even earlier exposures may have an impact on adult breast cancer risk [7]. Trichopoulos [8] suggested that the *in-utero* and perinatal period might be pathologically significant and that the risk of adult breast cancer could be related to high estrogen exposure in early life. There is also accumulating evidence that factors related to early exposure, such as birthweight, may be related to risk [9, 10].

There has been little research investigating possible effects of environmental exposures in early life on subsequent breast cancer risk. Using residence as a proxy measure for environmental exposures, we investigated whether there was any evidence of geographic clustering of adult breast cancer cases associated with their residences in early life. Clustering analyses have often been used to provide clues for the unknown etiology of disease, and thus to generate hypotheses for further epidemiologic research [11]. We looked at the geographic clustering of residence at early critical time points: at birth, at menarche, and at the woman's first birth. By comparing differences in clustering patterns between case and control residences, we were interested in identifying time periods critical to potential environmental exposures and subsequent breast cancer risk.

Methods

Population-based case–control study of breast cancer

We conducted a case–control study of breast cancer in western New York – the WEB study (Western New York Exposures and Breast Cancer Study) . Cases were women, age 35–79 with incident, primary, pathologically confirmed breast cancer diagnosed in Erie and Niagara counties during the period 1996–2001, with no previous cancer diagnosis other than non-melanoma skin cancer. Controls were frequency matched to cases on age, race, and county of current residence; controls under 65 years of age were randomly selected from a New York State Department of Motor Vehicles list and those 65 years and over were chosen from a Health Care Finance Administration list. We ascertained cases by having a nurse–case finder visit the pathology departments of almost all hospitals in these counties. One hospital which did not participate does almost no cancer surgery and refers patients to other participating hospitals. For the one other hospital that did not participate, breast cancer cases were identified in the practice of the breast surgeons who see more than 99% of the cases from that hospital. Extensive in-person interviews and self-administered questionnaires were used to ascertain lifetime residential history and other breast cancer risk factors. A total of 1166 cases and 2105 controls were interviewed. Response rates were 72 and 65% for cases and controls, respectively.

All participants were asked to complete a lifetime residential history, to list the street address, town/city and zip code for their current address and then all other previous addresses throughout their lifetime. Participants provided 20,240 addresses, an average of approximately six addresses for each individual. In this study we focused on residence at the time of the participants' birth, menarche, and at the time that she had her first birth. Analyses were restricted to women residing in Erie or Niagara counties at each of these time points. There were, of course, participants whose addresses were the same for two or more of these times.

For women with incomplete residential information, additional information was obtained using historical city directories. We used these directories to find old addresses, and utilized various resources, such as web searches and commercial address databases for recent addresses. We also examined validity and reliability of reports of earlier residences in a number of ways. For birth addresses, we asked for information on birth address twice and have collected information on reliability of response. For the other time periods, we used information on maiden name and partial address information provided by the participants to search for records in city directories for the appropriate time periods. To improve our ability to geocode addresses, we developed several strategies. First, all addresses were standardized to be matched with the standard format used in GIS. We used the enhanced version of TIGER (Topologically Integrated Geographic Encoding and Referencing Systems), GDT/Dynamap 2000 [12], and overall matching rates were improved about 15–20 % when compared with the use of TIGER as a reference theme. We also used the stand-alone address cleaner ZP4 (Semaphore Co.) to correct and update zip code information to be matched with United States Postal Services certified addresses.

More than 85% of addresses were geocoded using the above strategies and resources. We failed to geocode some addresses primarily because of missing residential information, such as missing street numbers or street names. Since we are dealing with historical residential information, the likelihood of missing previous residential information was higher than that for current residential information. Table 1 is a summary table showing the numbers of cases and controls with complete residential information who resided in the two counties for each of the time periods. The percentage of missing residential information associated with

Table 1. Residential history of breast cancer cases and controls: numbers and percentage of complete and missing residences in Erie and Niagara counties: WEB Study, 1996–2001

	Complete residence		Incomplete or missing residence		Total eligible Erie and Niagara county residence at each time period	
	Case	Control	Case	Control	Case	Control
Birth	505 (79.9%)	804 (81.0%)	127 (20.1%)	189 (19.0%)	632	993
Menarche	673 (87.3%)	1143 (88.1%)	98 (12.7%)	154 (11.9%)	771	1297
First birth	616 (86.4%)	1153 (87.3%)	97 (13.6%)	167 (12.7%)	713	1320

each early life event was highest for birth addresses, at about 20%.

Clustering analyses of residences

To compare clustering patterns of breast cancer cases and controls at each time period, the primary method used was based on the k-function [13]. The k-function for a point process is defined as the number of events within distance h of an arbitrary event, divided by the overall intensity of events. It is estimated by

$$\lambda\hat{k}(h) = \sum_{i=1}^{n}\sum_{j=1}^{n} w(s_i, s_j)^{-1} I(d_{ij} \leq h)/n, \quad h > 0$$

where n is the number of events, λ is the expected density of events in the study region, h is the pre-specified distance, d_{ij} is the Euclidian distance between point i and point j, I is an indicator function that is equal to one if inter-event distances (d_{ij}) are less than or equal to h, and zero otherwise, and $w(s_i, s_j)$ is an edge correction estimator which is the proportion of the circumference of a circle centered at s_i, passing through s_j and that is inside the study area A [14]. Under the null hypothesis of spatial randomness, the expected value of $k(h)$ is πh^2. Geographic clustering will yield values of the k-function that are greater than this, since clustering will result in more pairs of points separated by a distance of h than would be expected in a random pattern.

We used the difference between k-functions for cases and controls to compare two patterns (*i.e.*, $D(h) = k_{\text{case}}(h) - k_{\text{control}}(h)$). Positive values of $D(h)$ indicate spatial clustering of cases relative to the spatial clustering of controls. Under the null hypothesis of random labeling of cases and controls, the expected value of $D(h)$ is zero, indicating that the k-functions of the cases and controls are the same. The test statistic, $D(h)$, was calculated with confidence envelopes using the *splancs* library in *S-plus* [15]. We obtained the approximate 95% confidence limits for two standard errors ($\pm 2\sqrt{Var\{D(h)\}}$) at the $\alpha = .05$ level [16]. When the estimated function $D(h)$ deviated from zero by greater than two standard deviations, we interpreted this as a statistically significant difference between the case and control patterns.

We also employed a spatial clustering method to identify significant geographic clusters of breast cancer cases. The spatial scan statistic [17], which considers the likelihood of observing the actual number of cases inside of a circle under the null hypothesis of no clustering, was applied to residence at early life events. We were mainly interested in spatial clustering of high rates, and employed the Bernoulli model based on the locations of individual cases and controls [18]. In addition, odds ratios (OR) and 95% confidence intervals (95% CI) were obtained using logistic regression, adjusting for age, education, age at menarche, parity, history of benign breast disease, family history of breast cancer. All analyses were conducted for the entire group of study participants and for data stratified on menopausal status. Women were considered post-menopausal if their menses had ceased permanently and naturally. Among other women, participants were also considered post-menopausal if any of the following conditions were true: they were on hormone replacement therapy and were over age 55, they had had a bilateral oophorectomy, they had had a hysterectomy without removal of the ovaries and they were older than 50, their menses had ceased permanently due to radiation or other medical treatment and they were older than 55.

Results

Characteristics of subjects included in the analysis, subjects with missing residential information, and subjects excluded due to residence outside of Erie and Niagara counties, are shown in Table 2. About half of the sample was excluded for each time period; the highest percentage of ineligible cases and controls was at the birth residence (46 and 51% respectively). However, we found little difference in characteristics between

Table 2. Characteristics of subjects included in the analysis, subjects with missing residential information, and subjects excluded due to residence outside of the study area (Mean ± SD): WEB Study, 1996–2001

	Cases (n = 1166)			Controls (n = 2105)		
	Included	Missing	Ineligible*	Included	Missing	Ineligible*
Birth	(n = 505)	(n = 127)	(n = 534)	(n = 804)	(n = 189)	(n = 1112)
Age (years)	56.5 ± 10.9	60.0 ± 11.0	58.9 ± 11.3	55.6 ± 11.7	58.0 ± 11.8	59.4 ± 11.7
Education (years)	13.5 ± 2.4	13.1 ± 2.5	13.6 ± 2.7	13.4 ± 2.2	13.2 ± 2.2	13.3 ± 2.5
Parity	2.2 ± 1.5	2.4 ± 1.7	2.4 ± 1.8	2.6 ± 1.8	2.7 ± 1.8	2.8 ± 1.8
Age at menarche (years)	12.4 ± 1.5	12.6 ± 1.5	12.7 ± 1.7	12.7 ± 1.7	12.6 ± 1.6	12.7 ± 1.7
Age at first birth (years)	24.3 ± 4.6	23.5 ± 4.5	24.2 ± 5.1	24.5 ± 4.3	23.5 ± 4.2	24.0 ± 4.7
Pre-menopausal (%)	35.2	18.9	26.4	31.7	28.6	24.6
Body Mass Index	28.2 ± 6.4	28.4 ± 5.8	28.7 ± 6.4	28.0 ± 6.2	28.2 ± 6.0	28.4 ± 6.4
Family history of breast cancer (% yes)	21.3	18.9	20.2	12.7	16.2	12.4
History of benign breast disease (% yes)	34.9	37.0	32.8	22.3	25.9	20.6
Menarche	(n = 673)	(n = 98)	(n = 395)	(n = 1143)	(n = 154)	(n = 808)
Age (years)	56.6 ± 10.7	60.1 ± 11.6	59.5 ± 11.3	56.0 ± 11.7	60.2 ± 11.7	59.9 ± 11.6
Education (years)	13.5 ± 2.4	12.8 ± 2.6	13.6 ± 2.8	13.4 ± 2.2	13.0 ± 2.3	13.3 ± 2.6
Parity	2.2 ± 1.6	2.8 ± 1.8	2.5 ± 1.8	2.6 ± 1.8	2.9 ± 2.1	2.9 ± 1.8
Age at menarche (years)	12.5 ± 1.6	12.8 ± 1.5	12.7 ± 1.7	12.7 ± 1.6	12.6 ± 1.7	12.7 ± 1.7
Age at first birth (years)	24.3 ± 4.6	23.0 ± 4.3	24.2 ± 5.3	24.4 ± 4.5	23.8 ± 4.4	24.0 ± 4.6
Pre-menopausal (%)	30.3	24.5	24.6	33.8	23.4	23.3
Body Mass Index	28.1 ± 6.2	29.5 ± 6.4	28.7 ± 6.5	28.3 ± 6.5	27.6 ± 5.5	28.2 ± 6.1
Family history of breast cancer (% yes)	20.2	22.4	20.6	13.1	13.2	12.1
History of benign breast disease (% yes)	34.5	40.8	31.9	22.3	19.5	21.3
First Birth	(n = 616)	(n = 97)	(n = 453)	(n = 1153)	(n = 167)	(n = 785)
Age (years)	57.4 ± 11.1	58.9 ± 10.8	58.5 ± 11.2	57.0 ± 11.7	60.6 ± 10.7	58.5 ± 12.0
Education (years)	13.4 ± 2.3	13.0 ± 2.9	13.7 ± 2.8	13.3 ± 2.2	13.0 ± 2.1	13.4 ± 2.6
Parity	2.7 ± 1.3	3.1 ± 1.5	1.7 ± 1.9	3.0 ± 1.5	3.4 ± 1.7	2.2 ± 2.0
Age at menarche (years)	12.6 ± 1.5	12.5 ± 1.8	12.6 ± 1.6	12.7 ± 1.6	12.6 ± 1.5	12.7 ± 1.7
Age at first birth (years)	24.8 ± 4.8	22.2 ± 4.1	23.4 ± 4.9	24.7 ± 4.6	22.9 ± 3.5	23.3 ± 4.4
Pre-menopausal (%)	29.4	26.8	26.0	32.2	18.0	26.6
Body Mass Index	28.4 ± 6.3	30.1 ± 6.6	28.2 ± 6.3	28.1 ± 6.1	28.3 ± 6.5	28.4 ± 6.4
Family history of breast cancer (% yes)	21.2	23.7	18.6	11.5	19.4	13.6
History of benign breast disease (% yes)	34.7	37.1	32.7	21.0	25.1	22.0

* Ineligible due to residence outside of Erie and Niagara county.

those subjects included and those subjects with addresses outside of these two counties.

Mapping was used to identify geographic patterns of breast cancer cases and controls for each of the early life events. Maps showing the locations of cases and controls in Figure 1 portray the underlying geographic patterns of breast cancer cases and controls in the study area. The rectangular region was used instead of the actual county boundary as an approximate boundary of the study area to protect individuals' confidentiality. The purpose of such mapping is to inspect patterns visually – the first step in any spatial analysis. Geographic patterns do not appear to vary much from one time period to the next, and they appear to reflect patterns of population distribution in the study area. However, it is difficult to determine whether they were clustered or dispersed relative to population from visual inspection alone, because of the large number of data points.

To assess potential effects of geographic selection bias in our study, we also examined the distribution of current residence in relation to other population data on the geographic distribution of breast cancer cases and the general population. We did not find differences in the geographic distribution of participating and non-participating cases, or between controls and the underlying population, except some tendency for both cases and controls living closer to the interview site to be somewhat more likely to participate than those living further away.

Spatial clustering of residences associated with early life events

We obtained differences between the case and control patterns for locations associated with each early life event. The k-function differences for values of h up to 15 miles, with approximate 95% confidence envelopes, are shown

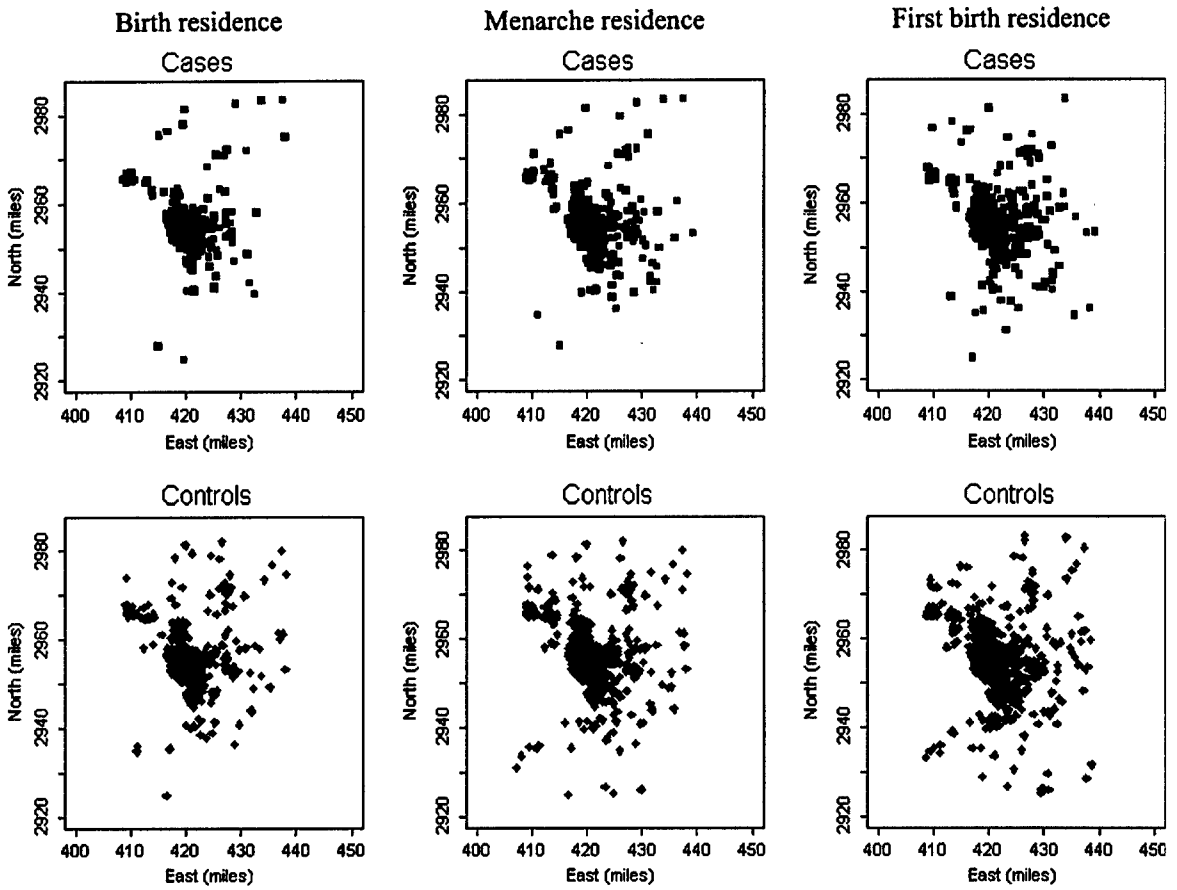

Fig. 1. Residential location of breast cancer cases and controls at each time period: WEB Study, 1996–2001.

in Figure 2. The maximum value of h is generally taken as one-third of the linear extent of the study area [19]. Any patterns beyond this scale can be disregarded, since either peaks or troughs in this geographic scale are difficult to interpret, and are potentially misleading. Figure 2a shows k-function differences for birth residence. It is clear that the estimated function shows strong evidence of spatial clustering, that is, of clustering of cases relative to controls. There was no significant difference up to three miles; statistically significant differences were detected beyond the scale of three miles. There is also evidence of some degree of clustering for breast cancer cases at menarche residence (Figure 2b). Estimates of the D-function are positive but not statistically significant up to seven miles; spatial clustering of breast cancer cases occurs at a scale of about 7–15 miles. For residence at women's first birth and for current residence, the difference is not statistically significant; the plot falls within the confidence interval over all distances (Figures 2c and d).

To determine whether there are any differences in clustering patterns by menopausal status, the k-function difference was performed for pre-menopausal and post-menopausal women separately (Figure 3). We found significant clustering of pre-menopausal breast cancer cases compared to controls for both birth and menarche residence (Figures 3a), while there is no evidence of clustering for post-menopausal breast cancer cases for either period (Figures 3b). We did not find evidence of clustering for first birth and current residence (at diagnosis) for either group (not shown). Estimated functions at birth residence show a strong clustering of pre-menopausal cases over the entire geographic scale with a peak at seven miles. Values are positive for post-menopausal cases, but not statistically significant. For menarche residence, we also observed a strong clustering of pre-menopausal cases with a peak at about 8–10 miles. Again differences are not statistically significant for post-menopausal women at menarche residence.

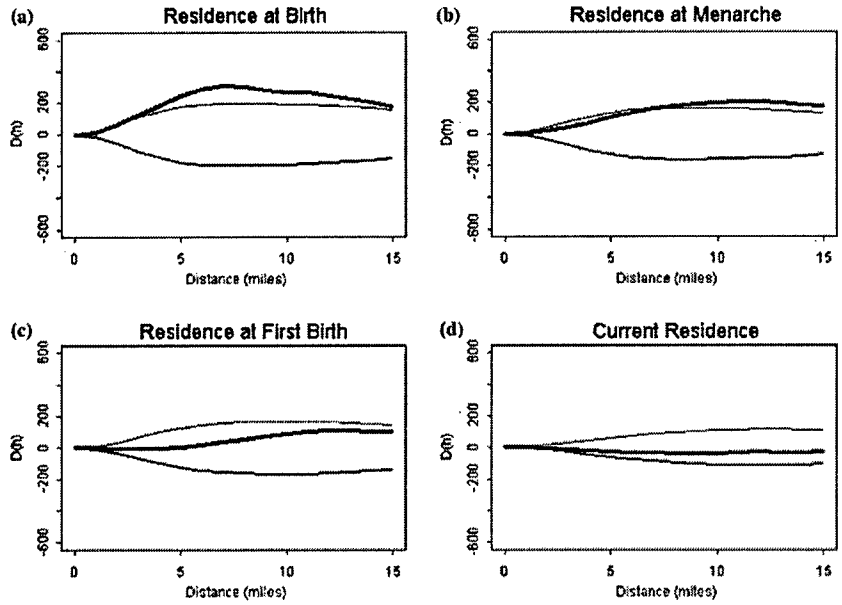

Fig. 2. k-function differences in clustering patterns between breast cancer cases and controls, WEB Study, 1996–2001: shown are k-function differences in black and 95% confidence limits in grey.

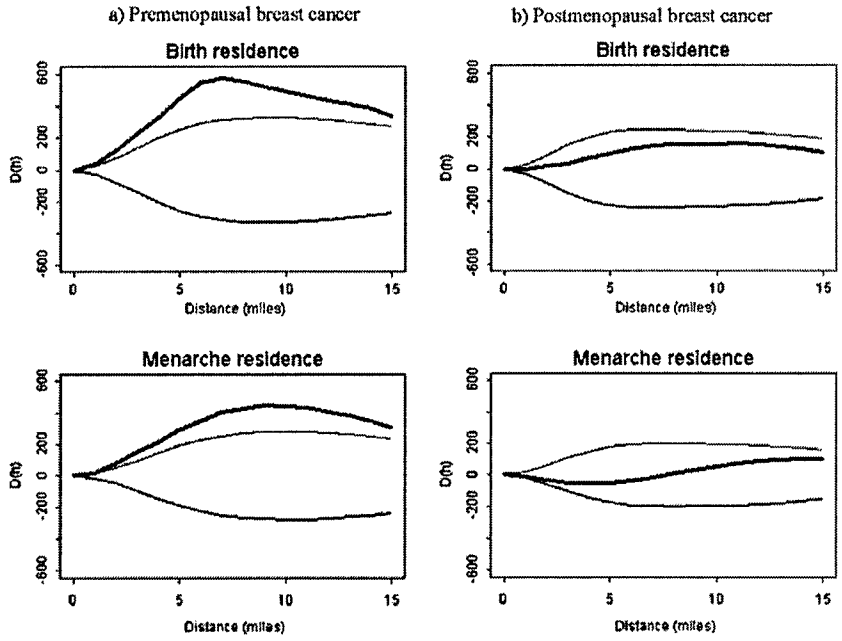

Fig. 3. k-function differences in clustering patterns between breast cancer cases and controls by menopausal status, WEB Study, 1996–2001.

Identifying the geographic location of breast cancer clusters

To identify the geographic location of areas with higher intensities for pre-menopausal cases in the study area, the spatial scan statistic was applied to residences of pre-menopausal women at the time of birth and menarche. Maps in Figure 4 present results of the clustering analysis. The circle in Figure 4a indicates

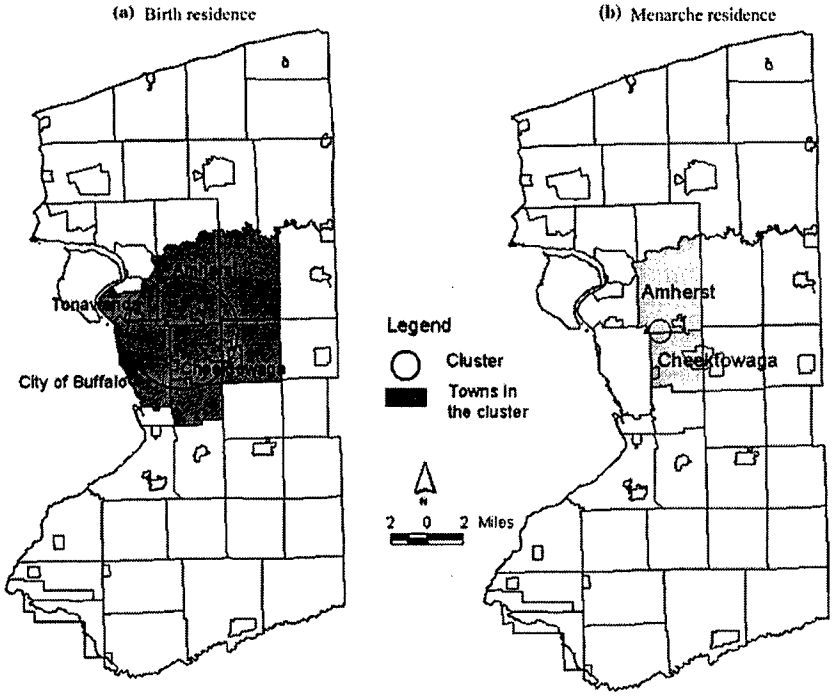

Fig. 4. Geographic clustering of residence at birth and menarche: pre-menopausal breast cancer, WEB Study, 1996–2001.

clustering of birth residence for pre-menopausal cases when compared to controls. We found a circular cluster of birth residence for breast cancer cases with a 5.7-mile radius in the area including part of the city of Buffalo, and the towns of Amherst, Cheektowaga, and Tonawanda (shaded areas). There are 100 observed breast cancer cases inside the cluster, while 76 breast cancer cases are expected. The cluster was significant at <0.01 with 999 Monte-Carlo simulations.

Further, we examined breast cancer risk associated with residence in the cluster at the time of birth. When we compared other breast cancer risk factors, such as age, education, and age at menarche, for the pre-menopausal breast cancer cases and controls whose birth residence was inside the cluster to those who lived outside of cluster, we did not find significant differences between the two groups (data not shown). We observed an elevated breast cancer risk for pre-menopausal women living in the cluster at the time of birth. With subjects living outside the cluster as a reference group, the adjusted odds ratio was 2.65 (95% CI 1.75–4.0) after controlling for age, education, age at menarche, parity, history of benign breast disease, and family history of breast cancer.

We also identified clustering of menarche residence for pre-menopausal women and obtained similar results as for birth residence. We were able to identify a small clustering of menarche residences for pre-menopausal breast cancer cases. A small cluster in the center of those four towns was detected (Figure 4b). It is a cluster with 0.8 mile radius and is statistically significant at $p < 0.05$. The cluster contains nine observed and 3.1 expected breast cancer cases, yielding a relative risk (ratio of observed to expected breast cancer cases) of 2.9. A secondary cluster was also detected near the city of Buffalo. It has a three-mile radius and relative risk of 1.38 with 65 observed and 47 expected breast cancer cases, but it is not statistically significant ($p = 0.38$).

Discussion

To our knowledge, no other studies have examined clustering of residential locations associated with cancer during early life: studies have examined clustering of residential locations at the time of diagnosis or death [20]. Critical time periods, including birth, menarche, and women's first pregnancy, as important early life and reproductive events in women's life, may play a substantial role in the risk of breast cancer. Under the hypothesis that there may be sensitive time periods in women's lives that will carry greater risk for exposure,

the essential question was whether cases were more clustered than the underlying population, as represented by the controls. We found that cases were more clustered than controls at the time of birth and menarche, and it was due to clustering of residence for pre-menopausal, but not for post-menopausal breast cancer. The evidence for clustering of residential locations at birth and menarche was stronger than evidence for clustering at the time of women's first birth or other time periods in adult life. Our findings suggest that there may be identifiable etiological processes linking exposure and breast cancer risk, especially for pre-menopausal women, and that early exposures may be of particular importance.

This study provided a unique opportunity to examine clustering of breast cancer cases and controls at various points during early life. The facts that the study area had a relatively stable population and about 40% of study participants were lifetime residents, made the results more reliable. The evidence that residence in early life was important in the geographical clustering of breast cancer cases may be of particular importance for understanding environmental determinants of breast cancer. These findings suggest the importance of early or lifetime exposure in relation to disease risk in adult life, and also the potential role of the effects of migration on exposures and disease risk. Although migration can have a serious effect on the detection of geographical differences in disease risk, it has not been adequately addressed in previous clustering analyses [21]. Further investigations are required to prove any relationship between geographic clustering of residence and breast cancer risk, and the effects of residential changes on exposures should be considered in these studies.

Our finding of clustering was restricted to pre-menopausal breast cancer. We stratified on menopausal status because of evidence that there were differences in risk factors for pre- and post-menopausal women [22]. The mechanism of the observed difference is not clear. It could be that early life exposures impact pre-menopausal more than post-menopausal disease because of greater temporal proximity. There is some evidence, though not consistent, that other early exposures may differ by menopausal status. For example, there are data suggesting that birthweight may be more associated with pre- than with post-menopausal breast cancer [9, 23].

The results should be interpreted cautiously due to the fact that there may be some artifacts of the analysis. First, it is important to note that spatial point patterns are complex to summarize in a single way [24]. For example, the use of cumulative scales in the application of the *k*-function method may influence the outcome [25]. In particular, clustering is more likely to be detected on a larger geographic scale, and it tends to show continuous patterns over several neighboring scales due to the fact that the geographical scales are cumulative. Further refinement of methods to summarize spatial point patterns may provide more reliable results, as well as more accurate estimates of disease risk.

Second, this study is limited to current residents in the study area because we focused on the residential environment of Erie and Niagara counties; participants residing outside of these two counties at the time of each early life event were not included. The existence of missing residential information and potential selection bias due to non-participation may influence the results. As noted, we found no difference in participation by residence for cases compared to controls. Further we would expect that our findings on the clustering of early-life residence would be less subject to potential geographic selection bias than would current residence. We found a greater degree of clustering for residence at early life than for current residential location.

Further, the fact that residence at birth and menarche were often the same made it difficult to differentiate associations for the two time periods. For 22% of cases and 35% of controls, the menarche residence was the same as their birth residence. While the observed tendencies may be related to environmental exposures, it is also possible that clustering of residence at the time of birth or menarche may be due to clustering of other socioeconomic or demographic factors. Evaluation of the contribution of socioeconomic status to clustering of residences at birth and menarche is of special interest. There may be other factors associated with residence not measured in this study. The findings are still of interest for further study in order to understand what those exposures might be. We are now investigating the relation between spatio-temporal clustering of residences and exposures to environmental compounds, such as PAHs and benzene, to provide epidemiologic evidence of this finding.

Since the publication of John Snow's [26] well-known cholera map for the city of London in the 19th century, the relationship between the environment and disease has been one of the major research themes in medical geography. Geographic perspectives are of great use in describing geographical patterns of diseases, generating hypotheses on disease etiology, monitoring high risk areas of disease incidence, and suggesting possible causal factors of particular disease [27, 28]. Our study demonstrated that these GIS-based clustering analyses provide effective ways to explore spatial–temporal patterns of clustering. The findings show consistent results; the cluster identified by spatial analyses

remained significant when traditional epidemiologic methods were used, and it was not explained by potential confounders. A recent study comparing 'traditional' epidemiological methods, GIS, and point pattern analysis for use in the spatially referenced public health data concluded that results complement, rather than contradict or duplicate each other [29].

In summary, this analysis of breast cancer clustering in space provides evidence of geographic clustering of pre-menopausal, but not post-menopausal, breast cancer cases at the time of birth and menarche, suggesting a possible influence of exogenous risk factors on breast cancer at these time points. While it is not clear from these data what caused this spatial clustering, it is provocative in providing evidence of the importance of this early period in breast carcinogenesis. Further investigations on genetic susceptibility may be of relevance to identify different effects on pre- and post-menopausal breast cancer. It will also be meaningful to see whether there is temporal clustering of early-life residences as well as spatial clustering. This type of study also needs to be replicated in other settings.

Acknowledgements

This work was supported in part by NIH Grants 1R01 ES09816-01, 1R21 CA87138-01, and U.S. Army Medical Research Grants DAMD17-03-1-0475, DAMD17-00-1-0417.

References

1. Sturgeon SR, Schairer CG, McAdams M, Brinton LA, Hoover RN (1995) Geographic variation in mortality from breast cancer among white women in the United States. *J Natl Cancer Inst* **87**: 1846–1853.
2. Lacey JV, Jr., Devesa SS, Brinton LA (2002) Recent trends in breast cancer incidence and mortality. *Environ Mol Mutagen* **39**: 82–88.
3. Wolff MS, Collman GW, Barrett C, Huff J (1996) Breast cancer and environmental risk factors: epidemiologic and experimental findings. *Annu Rev Pharmacol Toxicol* **36**: 573–596.
4. Laden F, Hunter DJ (1998) Environmental risk factors and female breast cancer. *Annu Rev Public Health* **19**: 101–123.
5. Kuh D, Ben-Shlomo Y (1997) A Life Course Approach to Chronic Disease Epidemiology: Tracing the Origins of Ill-health from Early to Later Life. Oxford: Oxford University Press.
6. Barker DJP (1992) *Fetal and Infant Origins of Adult Disease*. London: BMJ Publishing.
7. Colditz GA, Frazier AL (1995) Models of breast cancer show that risk is set by events of early life: prevention efforts must shift focus. *Cancer Epidemiol Biomarkers Prev* **4**: 567–571.
8. Trichopoulos D (1990) Hypothesis: does breast cancer originate *in-utero*? *Lancet* **335**: 939–940.
9. Potischman N, Troisi R (1999) *In-utero* and early life exposures in relation to risk of breast cancer. *Cancer Causes Control* **10**: 561–573.
10. Lumey LH (1998) Prenatal oestrogens and breast cancer. *Paediatr Perinat Epidemiol* **12**: 361–365.
11. Elliott P, Wakefield JC, Best NG, Briggs DJ (2000) *Spatial Epidemiology: Methods and Applications*. Oxford: Oxford University Press.
12. GDT (2001) GDT Dynamp/2000 Version 11.1, Geographic Data Technology, INC. USA.
13. Ripley BD (1981) *Spatial Statistics*. New York: Wiley.
14. Ripley BD (1977) Modelling spatial patterns. *J R Stat Soc Ser* B **39**: 172–212.
15. Rowlingson B, Diggle PJ (1993) SPLANCS: spatial point pattern analysis code in S-plus. *Comput Geosci* **19**: 627–655.
16. Diggle PJ, Chetwynd AD (1991) Second-order analysis of spatial clustering for inhomogeneous populations. *Biometrics* **47**: 1155–1163.
17. Kulldorff M, Nagarwalla N (1995) Spatial disease clusters: detection and inference. *Stat Med* **14**: 799–810.
18. Kulldorff M (1997) A spatial scan statistic. *Commun Stat-Theor M* **26**:1481–1496.
19. Bailey TC, Gatrell AC (1995) *Interactive Spatial Data Analysis*. Harlow: Longman.
20. Timander LM, McLafferty S (1998) Breast cancer in West Islip, NY: a spatial clustering analysis with covariates. *Soc Sci Med* **46**: 1623–1636.
21. Rogerson PA, Han D (2002) The effects of migration on the detection of geographic differences in disease risk. *Soc Sci Med* **55**: 1817–1828.
22. McPherson K, Steel CM, Dixon JM (2000) ABC of breast diseases: breast cancer epidemiology, risk factors, and genetics. *Br Med J* **321**: 624–628.
23. Stavola B, Hardy R, Kuh D, Silva I, Wadsworth M, Swerdlow A (2000) Birthweight, childhood growth and risk of breast cancer in a British cohort. *Br J Cancer* **83**: 964–968.
24. Cressie N (1993) *Statistics for Spatial Data*. New York: Wiley.
25. Reader S (2000) Using survival analysis to study spatial point patterns in geographical epidemiology. *Soc Sci Med* **50**: 985–1000.
26. Snow J (1855) On the mode of communication of cholera. In: Snow on cholera being a reprint of two papers by John Snow, M.D. 1936. New York: The Commonwealth Fund.
27. Jones K, Moon G (1987) *Health, Disease, and Society*. London: Routledge.
28. Meade MS, Florin JW, Gesler WM (1988) *Medical Geography*. New York: Guilford.
29. Dunn C, Kingham S, Rowlingson B et al. (2001) Analysing spatially referenced public health data: a comparison of three methodological approaches. *Health Place* **7**: 1–12.